

AMENDMENTS TO THE CLAIMS

Claim 1 (currently amended): A computer-implemented method for searching a collection of items, wherein each item in the collection has a set of properties, comprising the steps of:

- obtaining a query composed of a first set of one or more properties; and
- obtaining a result based on applying a distance function to the query and an item in the collection having a second set of one or more properties, wherein
 - obtaining a result includes determining a third set of properties common to the first set of one or more properties and the second set of one or more properties, and
 - the distance function determines a distance between the query and the item in the collection based on ~~the number~~ a frequency of occurrence of items in the collection that have are ~~associated with~~ all of the properties in the third set of properties, wherein a higher ~~number~~ frequency of occurrence of items in the collection having ~~associated with~~ all of the properties in the third set of properties indicates a greater distance between the query and the item in the collection and a lower ~~number~~ frequency of occurrence of items in the collection having ~~associated with~~ all of the properties in the third set of properties indicates a smaller distance between the query and the item in the collection; and
 - providing a representation of the result to a user.

Claim 2 (original): The method of claim 1, further including the step of associating each item in the collection with a set of properties.

Claim 3 (previously presented): The method of claim 1, wherein the step of obtaining a result includes identifying one or more result items whose distance from the query is within a first threshold.

Claim 4 (previously presented): The method of claim 3, wherein the step of obtaining a result includes ranking the one or more result items according to their distance from the query.

Claim 5 (original): The method of claim 3, wherein the threshold is defined as a number of result items.

Claim 6 (original): The method of claim 3, wherein the threshold is defined as a distance.

Claim 7 (original): The method of claim 1, further including the step of returning the result.

Claim 8 (original): The method of claim 1, wherein the step of obtaining a query includes the step of mapping a received query to a set of one or more properties.

Claim 9 (original): The method of claim 1, wherein one or more of the properties are binary.

Claim 10 (original): The method of claim 1, wherein one or more of the properties are related by a partial order, and wherein, if an item is associated with a property, then the item is also associated with all ancestors of that property in the partial order.

Claim 11 (previously presented): The method of claim 10, wherein one or more of the properties represent numerical values or ranges, and wherein the partial order reflects a set of containment relationships among the numerical values or ranges.

Claim 12 (original): The method of claim 1, wherein the properties are grouped into equivalence classes.

Claim 13 (original): The method of claim 12, further including the step of grouping the properties into equivalence classes using clustering.

Claim 14 (original): The method of claim 13, wherein each property has a set of subproperties, wherein the clustering is performed such that the distance between two properties in the collection is correlated to the number of properties in the collection that are associated with all of the subproperties common to both properties.

Claim 15 (original): The method of claim 1, wherein the query corresponds to a single item in the collection.

Claim 16 (original): The method of claim 1, wherein the query corresponds to a plurality of items in the collection.

Claim 17 (original): The method of claim 1, wherein the query is independent of the items in the collection.

Claim 18 (original): The method of claim 1, wherein the step of obtaining a result is constrained to a subcollection of the items in the collection.

Claim 19 (original): The method of claim 18, wherein the subcollection is specified as an expression of properties.

Claim 20 (original): The method of claim 19, wherein the expression includes a subset of the set of properties that compose the query.

Claim 21 (original): The method of claim 1, wherein the step of obtaining a query includes identifying certain properties to be ignored in the step of obtaining a result.

Claim 22 (original): The method of claim 1, wherein the distance function is applied explicitly.

Claim 23 (original): The method of claim 1, wherein the distance function is applied implicitly.

Claim 24 (original): The method of claim 23, wherein the step of obtaining a result includes the step of iterating a random walk process to select potential result items.

Claim 25 (original): The method of claim 24, wherein the step of obtaining a result includes ranking the potential result items by frequency and selecting the potential result items having higher frequencies.

Claim 26 (original): The method of claim 23, wherein the step of obtaining a result includes iterating through one or more subsets of the query and identifying items associated with the one or more subsets.

Claim 27 (original): The method of claim 26, wherein the one or more subsets are prioritized according to the number of items in the collection that have all of the properties in each subset and wherein iterating through one or more subsets of the query is continued until a first threshold is reached.

Claim 28 (original): The method of claim 1, wherein the step of obtaining a result includes applying a Euclidean distance function.

Claim 29 (original): The method of claim 28, wherein the step of obtaining a result includes merging a first result determined by applying the distance function and a second result determined by applying the Euclidean distance function.

Claim 30 (original): The method of claim 28, wherein the step of obtaining a result includes determining a first result by applying either the distance function or the Euclidean distance function and applying the other distance function to the first result.

Claim 31 (currently amended): A computer-implemented method for analyzing two sets of properties from a plurality of sets of properties of items in a collection being searched, each item in the collection having a set of properties in the plurality of sets of properties, the method comprising the steps of:

determining a set of properties common to the two sets of properties;

determining a frequency of occurrence ~~the number~~ of sets of properties, ~~in from~~ the plurality of sets of properties, that include all the ~~set of~~ common properties;

assessing the distance between the two sets of properties as a function of the frequency of occurrence ~~number~~ of sets of properties that include all the ~~set of~~ common properties, wherein a higher frequency of occurrence ~~number~~ of sets of properties that include all the ~~set of~~ common properties indicates a greater distance, and a lower frequency of occurrence ~~number~~ of sets of properties that include all the ~~set of~~ common properties indicates a smaller distance; and

providing a representation of the distance to a user.

Claim 32 (currently amended): A computer-implemented method for analyzing the relationship between two items in a collection of items, wherein each item in the collection is associated with a set of properties, comprising the steps of:

obtaining a set of properties with which the two items are commonly associated;
and

determining the degree of commonality between the two items as a function of ~~the number a frequency of occurrence~~ of items in the collection that ~~have are associated with~~ all of the properties with which the two items ~~are commonly associated~~ have in common, wherein a higher ~~number~~ frequency of occurrence of items in the collection having associated with all of the properties with which the two items have in common ~~are commonly associated~~ indicates a lesser degree of commonality and a lower ~~number~~ frequency of occurrence of items in the collection having associated with all of the properties with which the two items have in common ~~are commonly associated~~ indicates a greater degree of commonality; and

providing a representation of the degree of commonality to a user.

Claim 33 (currently amended): A computer program product, residing on a computer readable medium, for use in searching a collection of items, wherein each item in the collection is associated with a set of properties, the computer program product comprising instructions for causing a computer to:

receive a query composed of one or more properties; and

obtain a result based on applying a distance function to the query and an item in the collection having a second set of one or more properties,

wherein the distance function determines a third set of properties common to the first set of one or more properties and the second set of one or more properties, and determines a distance between the query and the item in the collection based on ~~the number a frequency of occurrence~~ of items in the collection that ~~have are associated with~~ all of the properties in the third set of properties, wherein a higher ~~number~~ frequency of occurrence of items in the collection having associated with all of the properties in the third set of properties indicates a greater distance between the query and the item in the collection and a lower ~~number~~ frequency of occurrence of items in the collection having associated with all of the properties in the third set of properties indicates a smaller distance between the query and the item in the collection;
and

provide a representation of the result to a user.

Claim 34 (original): The computer program product of claim 33, wherein the instructions cause the computer to obtain a result by identifying exactly the items whose distance from the query is within a threshold.

Claim 35 (original): The computer program product of claim 33, wherein the instructions cause the computer to obtain a result by identifying approximately the items whose distance from the query is within a threshold according to a heuristic.

Claim 36 (original): The computer program product of claim 35, wherein the heuristic permits a trade-off between the accuracy and the performance of a search.

Claim 37 (original): The computer program product of claim 35, wherein the heuristic includes the use of a random walk process.

Claim 38 (currently amended): A computer system for managing data records comprising:

an information retrieval subsystem that stores and retrieves data records, each data record being associated with a set of properties; and

a similarity search subsystem that receives similarity search queries and processes similarity search queries based on a distance function, a similarity search query being associated with a first set of properties,

wherein the distance function determines a distance between the query and a data record in the collection having a second set of properties based on determining a third set of properties common to the first set of properties and the second set of properties, and determining the ~~number~~ frequency of occurrence of data records in the collection that are associated with all of the properties in the third set of properties, wherein a higher ~~number~~ frequency of occurrence of data records associated with all of the properties in the third set of properties indicates a greater distance between the query and the item and a lower ~~number~~ frequency of occurrence of data records associated with all of the properties in the third set of properties indicates a smaller distance between the query and the item.

Claim 39 (original): The computer system of claim 38, further including a clustering subsystem that employs the distance function of the similarity search subsystem to construct a graph.

Claim 40 (cancelled).

Claim 41 (cancelled).